

PRECISION MEASUREMENT OF THE HIGGS BOSON MASS AND SEARCH FOR  
DILEPTON MASS RESONANCES IN  $H \rightarrow 4\ell$  DECAYS USING THE CMS DETECTOR AT  
THE LHC

By

JAKE ROSENZWEIG

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2022

© 2022 Jake Rosenzweig

I dedicate this to Jacob Myhre.

## ACKNOWLEDGEMENTS

Without so many two- and four-legged blessings along the way, I could never have made it to this point in my academic career. Thus, I give my infinite thanks to the following folks.

To my high energy physics mentors, Professors Andrey Korytov and Guenkah Mitselmakher, for granting me this one-of-a-kind opportunity to do real *science* at CERN.

To my wife, Suzanne Rosenzweig, for showing me that dreams do come true. To my mother and father, Vicki and John, who always reassured me that I could achieve anything I put my mind to. Sleep peacefully, Mom. To my siblings, Alex, Ryan, Devin, Jace, and Claudia who frequently and gently reminded me that there was life outside of grad school.

Aunt Rach, Uncle Yuri,

To my mentor, Sheldon Friedman, and his wife, Rita Friedman (Rosenzweig), who chose to invest in my success at a young age. I have only made it this far thanks to your undying encouragement, love, and optimism. To Sheldon's best friend, Dr. Bernard Khoury, whose reputation has helped pave my road.

To Dr. Filippo Errico for his Dr. Lucien Lo To Dr. Noah Steinberg To Darin Acosta, for spending hours of discussion on every t To the gentle gents who introduced me to the world of CMS, Brendan Regnery and Bhargav Joshi.

To my comrades for showing me what it takes to survive the core courses, Dr. Atool Divakarla, Dr. Brien O'Brendan, Dr. Donyell Guerrero, and Dr. Vladinar Martinez.

To the many students who tagged along in our "CMS Office Hours": Sean Kent, Jeremiah Anglin, Cris Caballeros, Ari Gonzalez, Evan Koenig, Nik Menendez, Neha Rawal, John Rötter. And to the many students who let us practice our spiels:

To my mentee, Matthew Dittrich, for

To my Polish roommates in Saint-Genis-Pouilly for showing me what home away from home feels like.

To the many moms who generously gave unconditional support during the darkest times and unconditional love during the brightest times: Silet Wiley, Margaret Sherrill, Dawn Hood, Cyndi Reilly-Rogers.

To my childhood best friends: Jish, Willis, Shane, Zac, Duck, and Marcus for their constant

clever competition which has shaped me into the determined man I am today.

Big Tree:

And finally to Existence itself for this unpredictable, unbelievable blip of an experience called life.

## TABLE OF CONTENTS

|   | <u>page</u> |
|---|-------------|
| ACKNOWLEDGEMENTS .....  | 4           |
| LIST OF TABLES.....   | 7           |
| LIST OF FIGURES.....  | 9           |
| CHAPTER   |             |
| 1    INTRODUCTION .....   | 10          |
| 2    THE CMS DETECTOR.....  | 11          |
| 2.1  The Silicon Tracker .....  | 14          |
| 2.1.1  The Pixel Detector.....  | 14          |
| 2.1.2  The Strip Detector .....   | 16          |
| 2.2  The Calorimeters .....   | 16          |
| 2.2.1  Electromagnetic Calorimeter.....   | 16          |
| 2.2.2  Hadron Calorimeter.....  | 19          |
| 2.3  The Solenoid and the Steel Return Yoke.....  | 21          |
| 3    HIGGS BOSON MASS MEASUREMENT IN THE $H \rightarrow ZZ^* \rightarrow 4\ell$ CHANNEL ..... | 24          |
| 3.1  Introduction .....   | 24          |
| 3.2  Analysis Overview.....   | 24          |
| 3.3  Data Sets and Simulated Samples.....   | 26          |
| 3.3.1  Data Sets.....   | 26          |
| 3.3.2  Simulated Samples .....  | 26          |
| 3.4  Object Reconstruction.....   | 26          |
| 3.4.1  Electrons.....   | 26          |
| 3.4.2  Muons .....  | 26          |
| 3.4.3  Isolation .....  | 26          |
| 3.4.4  Final State Radiation.....   | 26          |
| 3.5  Event Selection .....  | 26          |
| 3.5.1  Triggers .....   | 26          |
| 3.5.2 $ZZ$ Candidate Selection .....  | 26          |
| 3.6  Background estimation.....   | 26          |
| 3.6.1  Irreducible background .....   | 26          |
| 3.6.2  Reducible background .....   | 27          |
| 4    SUMMARY.....   | 40          |
| REFERENCES .....  | 40          |

## LIST OF TABLES

Tables

page

## LIST OF FIGURES

| <u>Figures</u>  | <u>page</u> |
|---|-------------|
| 2-1 Life-size poster of the CMS detector, taken during CERN Open Days 2019 in the SX5 warehouse where parts of CMS were assembled. ....   | 11          |
| 2-2 Points 1 through 8 along the LHC. Collisions occur at Points 1 (ATLAS), 2 (ALICE), 5 (CMS), and 8 (LHCb), whereas the remaining points are used for LHC beam maintenance and testing. ....  | 12          |
| 2-3 Cut out of the CMS detector showing its various subdetector components. ....  | 13          |
| 2-4 A transverse view of CMS showing the “filtration process” as different particles pass through different subdetectors. A positron (solid red line) curves due to the presence of the magnetic field and gets stopped in the ECAL, creating an EM shower. A photon (blue dashed line) does not get detected at all by the Silicon Tracker, since it has no electric charge. It continues through to the ECAL and makes a shower here, like the positron. Charged hadrons (solid green line) will show curved tracks from the Silicon Tracker, may leave some trace in the ECAL, but primarily get stopped by the HCAL creating hadronic showers. Neutral hadrons (dashed green line) do not interact with the tracker, and only undergo EM showers a little in the ECAL, but show most energy deposits in the HCAL. Muons (solid blue line) are detected by the Silicon Tracker and then mostly pass through the other subdetectors without interacting until they finally reach the Muon System. Using the Lorentz force law and knowing which direction the magnetic field is pointing, one can deduce the sign of the charge of the particle. Based on the radius of curvature from the trajectory, one can then calculate the momentum and energy of the particle. .... | 15          |
| 2-5 (Left) A simulation of the silicon tracker, showing the 3 cylindrical layers of the pixel detector (pink), 4 layers of the TIB (green), and the 6 layers of the TOB (blue) of the strip detector. The endcap components are also shown. (Right) A picture of the real silicon tracker at the center of CMS. ....  | 16          |
| 2-6 A transverse view of the silicon pixel and strip detectors, explicitly labelling the different layers involved. ....  | 17          |
| 2-7 (Left) Cross sectional view of the electromagnetic calorimeter of CMS. (Right) One of the Dees which comprise the EE. Each square of $5 \times 5$ crystals constitutes a “supercrystal”. Figure taken from Ref. [1]. ....   | 18          |
| 2-8 (Left) ECAL crystals made from PbWO <sub>4</sub> are grown in a lab. (Right) Although made mostly of metal, ECAL crystals are transparent and have a photomultiplier detector attached at the end. ....   | 18          |
| 2-9 A cross-sectional quadrant of CMS showing the locations of the HCAL components: the barrel (HB), outer (HO), endcap (HE), and forward (HF) detectors. ....  | 21          |
| 2-10 A longitudinal cross section of CMS showing the values of the magnetic field over the volume of CMS and various field lines. The magnetic field reaches its maximum of 3.8 T in the center of the detector. ....   | 22          |

|   |    |
|---|----|
| 3-1 Events from 2016 UL data that pass 2P2F CR selections (black markers) are compared to the stacked 2P2F distributions of simulated samples ( $Z + \text{jets}$ , $t\bar{t} + \text{jets}$ , WZ, $Z/\gamma^*$ , ZZ). The results are split into the $4\ell$ final states: $4\mu$ (top left), $4e$ (top right), $2e2\mu$ (bottom left), $2\mu2e$ (bottom right). ....  | 32 |
| 3-2 Events from 2017 UL data that pass 2P2F CR selections (black markers) are compared to the stacked 2P2F distributions of simulated samples ( $Z + \text{jets}$ , $t\bar{t} + \text{jets}$ , WZ, $Z/\gamma^*$ , ZZ). The results are split into the $4\ell$ final states: $4\mu$ (top left), $4e$ (top right), $2e2\mu$ (bottom left), $2\mu2e$ (bottom right). ....  | 33 |
| 3-3 Events from 2018 UL data that pass 2P2F CR selections (black markers) are compared to the stacked 2P2F distributions of simulated samples ( $Z + \text{jets}$ , $t\bar{t} + \text{jets}$ , WZ, $Z/\gamma^*$ , ZZ). The results are split into the $4\ell$ final states: $4\mu$ (top left), $4e$ (top right), $2e2\mu$ (bottom left), $2\mu2e$ (bottom right). ....  | 34 |
| 3-4 Events from 2016 UL data that pass 3P1F CR selections (black markers) are compared to the stacked 3P1F distributions of simulated samples ( $Z + \text{jets}$ , $t\bar{t} + \text{jets}$ , WZ, $Z/\gamma^*$ , ZZ) and to the predicted contribution of 2-prompt-2-nonprompt-lepton processes to the 3P1F CR (red line). This prediction is obtained by making a distribution of all event weights given by Eq. 3-4 and stacking that on top of the WZ and ZZ distributions. The results are split into the $4\ell$ final states: $4\mu$ (top left), $4e$ (top right), $2e2\mu$ (bottom left), $2\mu2e$ (bottom right). .... | 35 |
| 3-5 Events from 2017 UL data that pass 3P1F CR selections (black markers) are compared to the stacked 3P1F distributions of simulated samples ( $Z + \text{jets}$ , $t\bar{t} + \text{jets}$ , WZ, $Z/\gamma^*$ , ZZ) and to the predicted contribution of 2-prompt-2-nonprompt-lepton processes to the 3P1F CR (red line). This prediction is obtained by making a distribution of all event weights given by Eq. 3-4 and stacking that on top of the WZ and ZZ distributions. The results are split into the $4\ell$ final states: $4\mu$ (top left), $4e$ (top right), $2e2\mu$ (bottom left), $2\mu2e$ (bottom right). .... | 36 |
| 3-6 Events from 2018 UL data that pass 3P1F CR selections (black markers) are compared to the stacked 3P1F distributions of simulated samples ( $Z + \text{jets}$ , $t\bar{t} + \text{jets}$ , WZ, $Z/\gamma^*$ , ZZ) and to the predicted contribution of 2-prompt-2-nonprompt-lepton processes to the 3P1F CR (red line). This prediction is obtained by making a distribution of all event weights given by Eq. 3-4 and stacking that on top of the WZ and ZZ distributions. The results are split into the $4\ell$ final states: $4\mu$ (top left), $4e$ (top right), $2e2\mu$ (bottom left), $2\mu2e$ (bottom right). .... | 37 |
| 3-7 Electron (top row) and muon (bottom row) fake rates evaluated using the OS Method vs. the $p_T$ of the probe lepton for 2016, 2017, and 2018 UL data (left, middle, right columns, respectively). ....  | 39 |

## CHAPTER 1 INTRODUCTION

The universe, while overwhelmingly vast, is comprised of a curiously small number of elementary particles. These particles and their strong, weak, and electromagnetic interactions with each other are accurately described by the Standard Model (SM). A major shortcoming of the SM was its inability to predict the masses of these particles.

This dissertation presents a precision measurement of the Higgs boson mass and using LHC proton-proton collision data from Run 2 data set from

The SM was not able to predict the masses of these particles until 1964 when the Brout-Englert-Higgs mechanism suggested that It wasn't until 1964 that the Brout-Englert-Higgs mechanism gave a self-consistent way to : by breaking the electroweak gauge symmetry of the vacuum would give rise to non-zero masses of the weak gauge bosons. This would yield a secondary effect too: there should exist a fundamental scalar boson which is the quantum of the so-called "Higgs field". On July 4th, 2012, this Higgs boson was discovered.

At first glance, the universe appears to be an overwhelmingly vast and complicated place. However upon closer inspection, it is comprised of only a few different kinds of fundamental particles. Particle physics has given rise to the Standard Model (SM) which mathematically describes these constituents and their interactions with each other.

The Standard Model (SM) is an impressively accurate mathematical theory which describes the fundamental particles of the universe and the rules for their possible interactions. Problematically though, the SM predicts that all particles are massless.

Get to the Higgs boson.

Why is it important? Knowing the mass of the Higgs boson

## CHAPTER 2

### THE CMS DETECTOR



Figure 2-1. Life-size poster of the CMS detector, taken during CERN Open Days 2019 in the SX5 warehouse where parts of CMS were assembled.

Weighing in at 14,000 tonnes, standing 5 stories tall (15 m), and reaching 29 m long, the Compact Muon Solenoid (CMS) experiment is one of two general-purpose particle detectors at the LHC (Fig. 2-1). CMS is situated approximately 100 m under the earth at the fifth collision point (Point 5) along the LHC (Fig. 2-2). In 2012, both CMS and its competing experiment, ATLAS, independently discovered the Higgs boson.

As discussed in Section (TODO: REF), the LHC collides bunches of protons every 25 ns to produce thousands of new particles which then travel away from the interaction point. CMS is built around the interaction point in a series of cylindrical subdetectors for nearly hermetic coverage so that most of the particles must travel through CMS. The detector sports a solenoid, after which CMS was named, which generates a 3.8 T uniform magnetic field that points longitudinally down the central axis of CMS. This strong magnetic field applies a Lorentz force on the outgoing charged particles, causing them to follow helical, momentum-dependent trajectories. These curved tracks are then better separated from one another which assists in particle identification. Neutral particles experience no Lorentz force and thus travel in straight lines.

The subdetectors measure the properties of the outgoing particles and carefully filter them out in a clever way (Fig. 2-3). Particles interact with the subdetectors, leaving so called “hits”

where they passed through. Hits are reconstructed into tracks. From the track curvature, deduce charge and momentum of the particles. Depending on which subdetector (or combination of subdetectors) was hit by the outgoing particles, the type of particle can be deduced. A few

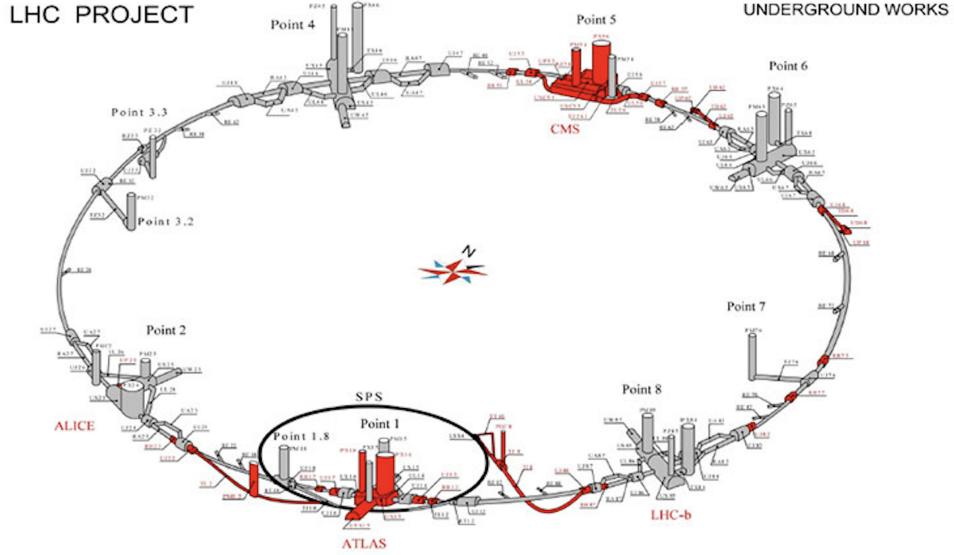


Figure 2-2. Points 1 through 8 along the LHC. Collisions occur at Points 1 (ATLAS), 2 (ALICE), 5 (CMS), and 8 (LHCb), whereas the remaining points are used for LHC beam maintenance and testing.

example particles and their associated tracks are shown in Fig. 2-4.

Before discussing each subdetector in the following sections, it is useful to define the coordinate system used in CMS: a typical, right-handed, three-dimensional Cartesian coordinate system ( $x, y, z$ ) is used, whose center  $(0, 0, 0)$  is placed at the nominal pp collision point within CMS. The  $x$ -axis points towards the center of the LHC, the  $y$ -axis points vertically upward, and the  $z$ -axis points westward towards the Jura mountains, tangential to the beam direction. Since CMS covers almost the entire spherical  $4\pi$  steradians around the interaction point, it is convenient to use spherical coordinates  $(r, \phi, \theta)$ , in which  $r$  measures the radial distance in the  $x$ - $y$  plane,  $\phi$  measures the azimuthal angle in the  $x$ - $y$  plane as measured from the  $x$ -axis, and  $\theta$  measures the polar angle as measured from the  $z$ -axis. When dealing with ultra-relativistic particles like those produced at the LHC, special relativistic effects like length contraction must be taken into account and so the coordinate  $\theta$  becomes frame-dependent. It is thus helpful to convert  $\theta$  to the

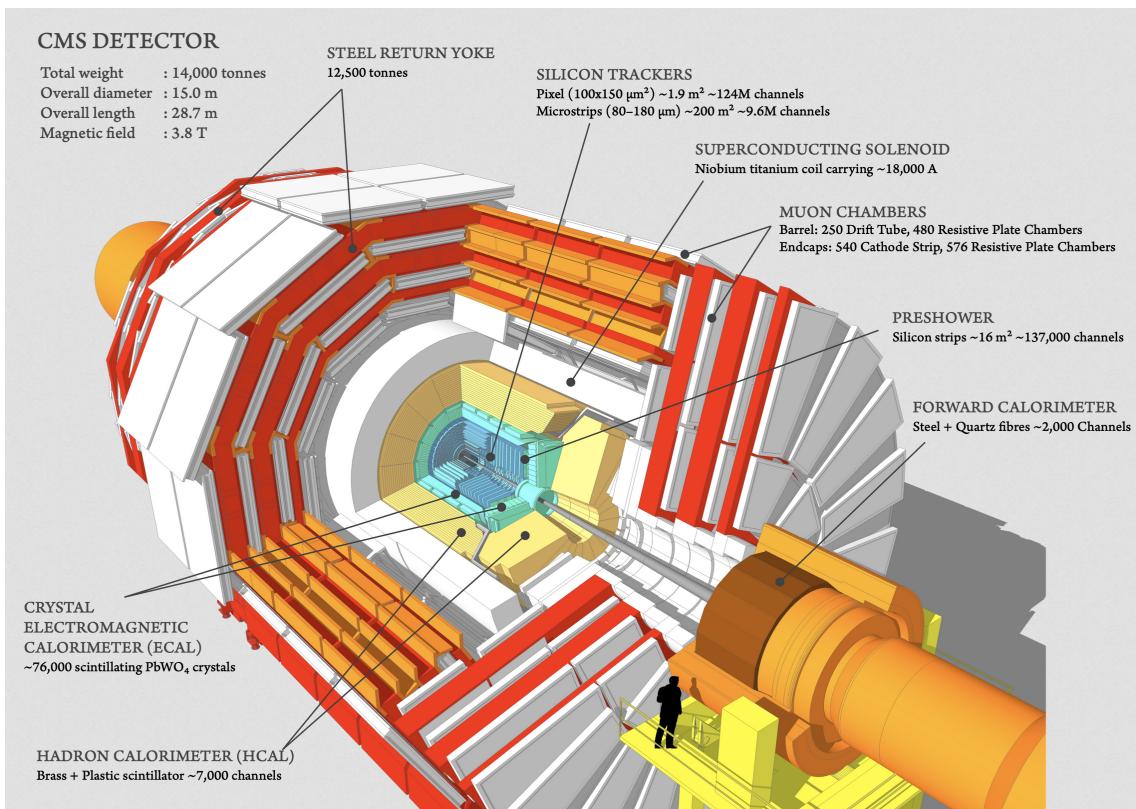


Figure 2-3. Cut out of the CMS detector showing its various subdetector components.

Lorentz-invariant quantity called pseudorapidity ( $\eta$ ), defined as  $\eta = -\ln[\tan(\theta/2)]$ .

## 2.1 The Silicon Tracker

At the heart of CMS is one of the world’s largest silicon detectors: the silicon tracker. The main goal of the silicon tracker is not to capture outgoing particles but to very precisely measure the hits from the charged particles as they pass through it. The tracker also assists in vertex identification, differentiating between primary and secondary vertices, the latter of which often comes from B meson decays. When multiple pp collisions occur within the same BX (pile up), the tracker distinguishes between proton collisions with a resolution of about  $100 \mu\text{m}$  longitudinally and  $50 \mu\text{m}$  transverse to the beam pipe. This is crucial to resolve which outgoing particles came from which pp vertex.

The tracker consists of two types of pure silicon detectors: the pixel detector and the strip detector, each of which is described in detail below.

### 2.1.1 The Pixel Detector

The innermost part of the silicon tracker is the pixel detector, which is the closest subdetector to the interaction point. The pixel detector is composed of 66 million silicon “pixels”, as shown in Fig. 2-5 (Left, pink). A single pixel is  $100 \mu\text{m} \times 150 \mu\text{m}$  and, collectively, they cover a sensitive area of  $1.9 \text{ m}^2$ . Because it sits only 8 cm away from the beam pipe, the pixel detector receives the highest particle flux than any other subdetector: around 10 million particles per  $\text{cm}^2$  per second.

The pixel detector is made of three cylindrical layers and two endcaps that surround the beam pipe. In total, the pixel detector has around 6,000 connections (channels) per  $\text{cm}^2$ .

After the LHC Run 1 was completed, the accelerator received luminosity upgrades during the 2013-2014 long shutdown period. To handle these higher luminosities, the pixel detector was replaced by the CMS Phase-1 pixel detector during the LHC technical stop in 2016-2017. The upgrades outfitted the detector with four barrel layers and three endcap disks per side, which allowed for particle detection up to  $|\eta| < 2.5$ . The overall mass of the pixel detector decreased and granted the detector with better tracking capability.

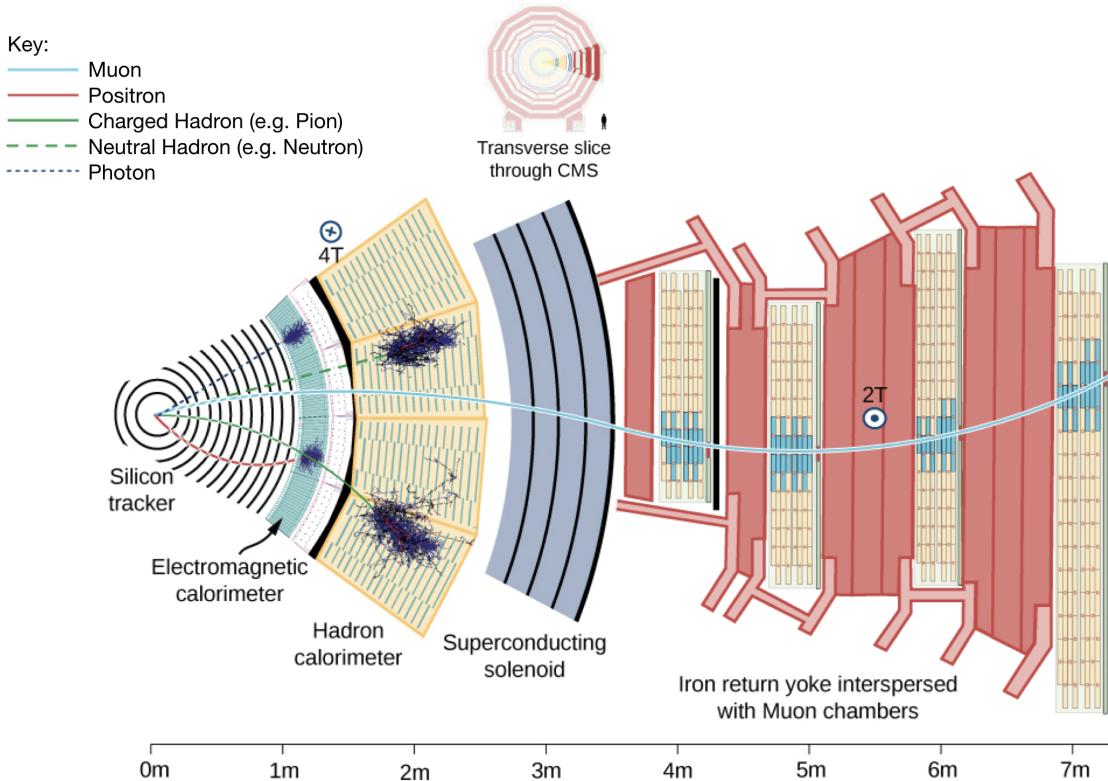


Figure 2-4. A transverse view of CMS showing the “filtration process” as different particles pass through different subdetectors. A positron (solid red line) curves due to the presence of the magnetic field and gets stopped in the ECAL, creating an EM shower. A photon (blue dashed line) does not get detected at all by the Silicon Tracker, since it has no electric charge. It continues through to the ECAL and makes a shower here, like the positron. Charged hadrons (solid green line) will show curved tracks from the Silicon Tracker, may leave some trace in the ECAL, but primarily get stopped by the HCAL creating hadronic showers. Neutral hadrons (dashed green line) do not interact with the tracker, and only undergo EM showers a little in the ECAL, but show most energy deposits in the HCAL. Muons (solid blue line) are detected by the Silicon Tracker and then mostly pass through the other subdetectors without interacting until they finally reach the Muon System. Using the Lorentz force law and knowing which direction the magnetic field is pointing, one can deduce the sign of the charge of the particle. Based on the radius of curvature from the trajectory, one can then calculate the momentum and energy of the particle.

### 2.1.2 The Strip Detector

The outer part of the silicon tracker is called the strip detector, which has 10 million detector strips spread across 10 cylindrical layers. The first 4 layers belong to the tracker inner barrel (TIB) and the remaining 6 layers belong to the tracker outer barrel (TOB), Fig. 2-5 (Left, green and blue, respectively). Both the TIB and TOB have two endcaps associated with them, the TID and TEC, respectively. Accounting for all of its components, the strip detector is sensitive to  $200 \text{ m}^2$ .

Fig. 2-6 gives a clearly-labelled transverse illustration of the pixel and strip detectors.

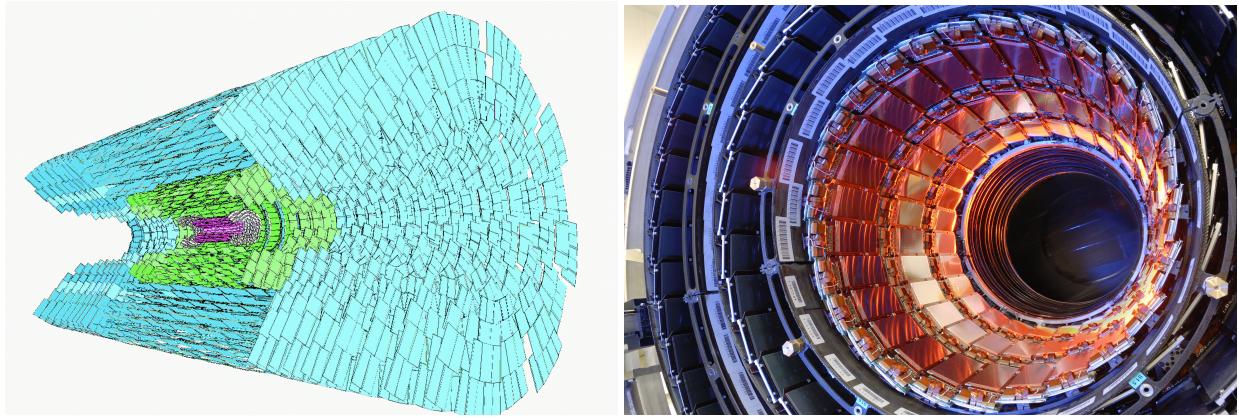


Figure 2-5. (Left) A simulation of the silicon tracker, showing the 3 cylindrical layers of the pixel detector (pink), 4 layers of the TIB (green), and the 6 layers of the TOB (blue) of the strip detector. The endcap components are also shown. (Right) A picture of the real silicon tracker at the center of CMS.

## 2.2 The Calorimeters

### 2.2.1 Electromagnetic Calorimeter

**Overview:** Particles that pass through the silicon tracker 2.1 encounter the electromagnetic calorimeter (ECAL). Those particles which interact electromagnetically but not strongly, mostly photons and electrons, are typically absorbed by the ECAL. The particle's energy is then transferred to the ECAL in the form of an electromagnetic (EM) shower. The size and shape of the EM shower provide information about the particle's energy and trajectory. Since the Higgs boson can decay into two photons, it was essential that the ECAL was able to detect this decay mode.

**Design:** The ECAL is a hermetic, cylindrical, homogeneous sub-detector that consists of a barrel (EB), two endcaps (EE), and a preshower detector in front of each endcap (Figure 2-7,

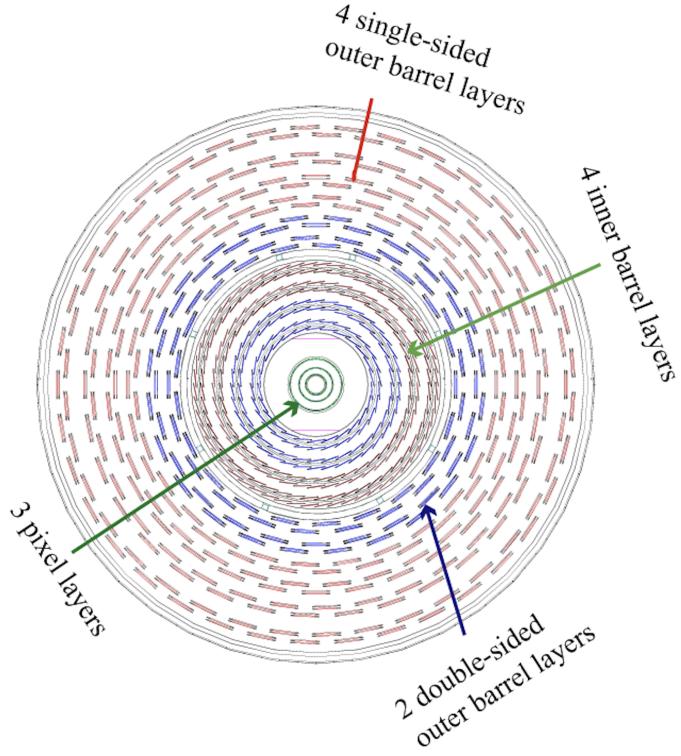


Figure 2-6. A transverse view of the silicon pixel and strip detectors, explicitly labelling the different layers involved.

Left). The EB covers  $|\eta| < 1.479$  while the EE covers  $1.479 < |\eta| < 3.0$ . The entire subdetector is composed of transparent lead tungstate ( $\text{PbWO}_4$ ) crystals that point axially towards the center of CMS (the interaction point). The transparent crystals, one of which is shown in Fig. 2-8 (Left), have a high density ( $8.28 \text{ g/cm}^3$ ) which provides the ECAL with radiation resistance and a short radiation length ( $X_0 = 0.89 \text{ cm}$ ). Because so many crystals are used (61,200 crystals in the EB and 7,324 in the EE), the ECAL has excellent energy resolution and fine granularity. Each endcap is composed of two “Dee”s, one of which is shown in Figure 2-7 (Right). A single Dee carries 3,662 crystals. Crystals in the barrel are tapered, having front face dimensions  $2.2 \times 2.2 \text{ cm}^2$ , back face dimensions  $2.6 \times 2.6 \text{ cm}^2$ , and are 23.0 cm long ( $25.8 X_0$ ). Crystals in the endcaps are also tapered, with front face dimensions  $2.862 \times 2.862 \text{ cm}^2$ , back face dimensions  $3.0 \times 3.0 \text{ cm}^2$ , and are 22.0 cm long ( $24.7 X_0$ ). This gives a single crystal from the barrel a volume of approximately  $132.5 \text{ cm}^3$  (mL), about the volume of a small cup of coffee, yet it weighs 1.5 Kg. (REF:PDG).

REF:PDG Particle Data Group collaboration, S. Eidelman et al., *Review of particle physics*,

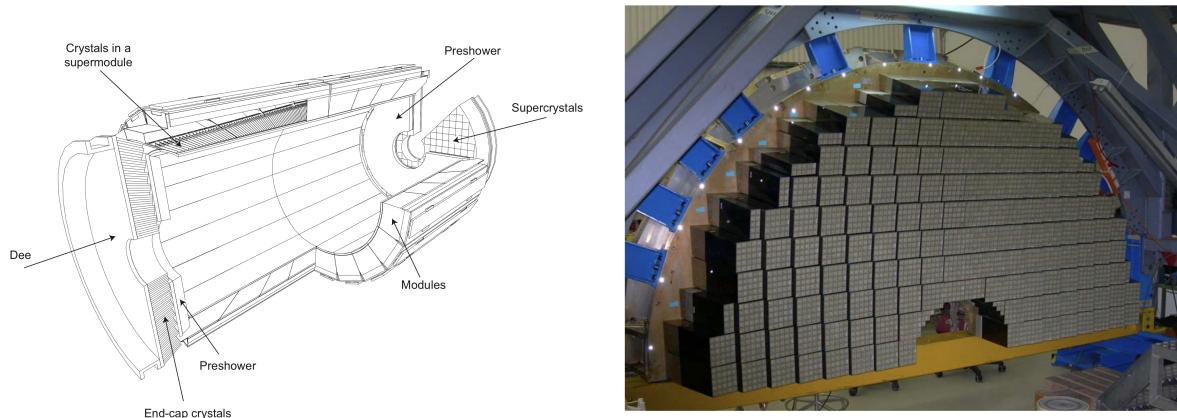


Figure 2-7. (Left) Cross sectional view of the electromagnetic calorimeter of CMS. (Right) One of the Dees which comprise the EE. Each square of  $5 \times 5$  crystals constitutes a “super-crystal”. Figure taken from Ref. [1].

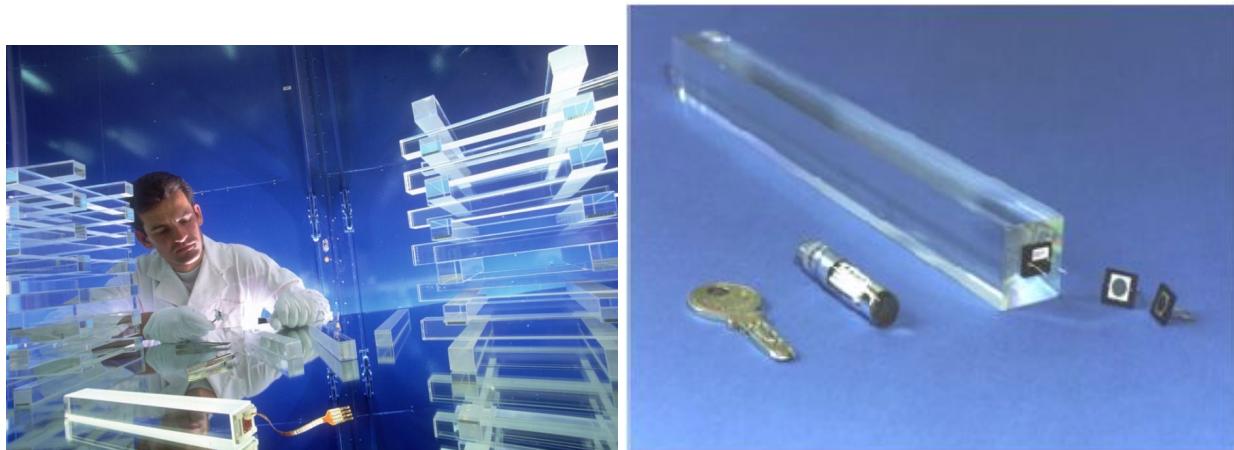


Figure 2-8. (Left) ECAL crystals made from PbWO<sub>4</sub> are grown in a lab. (Right) Although made mostly of metal, ECAL crystals are transparent and have a photomultiplier detector attached at the end.

**Physics:** When electrons or photons pass through the ECAL, they create an EM shower. Electrons radiate more photons as they accelerate around PbWO<sub>4</sub> nuclei, in a process called *bremsstrahlung*. Meanwhile, near the presence of a nucleus, high-energy photons pair produce into e<sup>+</sup>e<sup>-</sup>. This cycle of electron and photon production disperses the initial particle energy into a spray of lower- and lower-energy particles; an EM shower (REFERENCE EM SHOWER FIG).

SHOW PICTURE OF EM SHOWER.

The ECAL crystals then scintillate (emits photons) in proportion to the amount of energy deposited by the interacting particle. The scintillator photons are detected by avalanche photodiodes on the back of each barrel crystal or by vacuum phototriodes in the endcap crystals (Fig. 2-8, Right). Conveniently after 1 bunch crossing (25 ns), Approximately 80% of the scintillated light is emitted.

An energy deposit in the ECAL could come from either an electron or a photon. In order to tell the difference, information from the silicon tracker is used. Charged particles, like electrons, will leave hits in the tracker and follow a curved path, whereas photons are electrically neutral and thus will not show any signs within the silicon tracker. So long as the tracker and ECAL communicate effectively with each other, then they help distinguish between electrons and photons. Charged hadrons interact only minimally with the ECAL, instead continuing on to the Hadron Calorimeter. Neutral hadrons can be detected by the ECAL preshower near the ECAL endcaps which helps distinguish a single photon from  $\pi^0$  mesons as they decay into two photons with a narrow opening angle, making it look as if the two photons are a single photon. The preshower detector allows CMS to distinguish between low-energy diphoton pairs and single high-energy photons.

NEED SMOOTH TRANSITION INTO HCAL. What about those hadrons? They got through the ECAL... To detect hadrons effectively, we need a Hadron Calorimeter.

## 2.2.2 Hadron Calorimeter

**Overview:** The particles that survive the ECAL, typically only muons and hadrons, then enter the hadron calorimeter (HCAL). Its primary purpose is to absorb the hadronic matter emerging from the interaction point and measure the corresponding jet energies. The absorbed jets cause the HCAL to scintillate photons which are then converted into electrical signals. These signals are used to deduce the original jet energies and any missing transverse energy ( $E_T^{\text{miss}}$ ) from the event.

**Design:** Dissimilar to the ECAL (subsec. 2.2.1) in material composition but similar in shape, the HCAL is a brass cylindrical scintillator. Although it has a barrel (HB) and two endcaps

(HE), it has two more detectors than the ECAL: the outer calorimeter (HO) and the forward calorimeter (HF). The HB spans the pseudorapidity range  $|\eta| < 1.3$ , the HE spans  $1.3 < |\eta| < 3$ , and the HF spans  $3 < |\eta| < 5.2$ , as shown in Figure 2-9. With a thickness of over 1 m, the HB is sandwiched between the barrels of the ECAL and the solenoid (subsec. 2.3) at radial values  $r = 1.77$  m and  $r = 2.95$  m, respectively. Because the HB and HE are located within the solenoid's strong magnetic field of 3.8 T, they both were both constructed out of a non-magnetic absorber called *C26000 cartridge brass*. This absorber has a density of  $8.53 \text{ g/cm}^3$  and an interaction length ( $\lambda_I$ ) of 16.42 cm. The thickness of the HB increases as  $1/\sin\theta$  so that at  $|\eta| = 0(1.3)$  the absorber thickness is  $5.82(10.6) \lambda_I$ . The HB is composed of two half-barrels, where each half-barrel is built from 18 identical azimuthal wedges and each wedge spans  $20^\circ$ . Each wedge is divided into four  $\phi$  segments so that a single  $\phi$  segment spans  $\Delta\phi = 0.087$ .

Since the volume available to the HCAL is so limited, and in order to stop any particles that might traverse the entire HCAL and solenoid, the HO (the “tail catcher”) is situated outside the barrel of the solenoid. The HF is located 11.2 m from the interaction point.

All tiles within a single  $\phi$  segment are grouped together into a single tray unit. The scintillator is also segmented into 16  $\eta$  sectors, the first(last) of which is located at  $|\eta| = 0(1.3)$ . This way each tile covers  $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$ . Each layer has 108 trays.

**Physics:** Since hadrons are the only particles to interact via the strong force, the HCAL is designed to have a high nuclear density. This ensures ample opportunity for hadrons to radiate gluons and convert with the Similal to the ECAL, the HCAL will scintillate in proportion to the amount of energy of the captured particle. The incoming hadrons will *hadronize* (i.e., produce a hadronic shower), generating jets of quarks and gluons which are bound in various ways forming protons, neutrons, pions, kaons, etc. Interestingly, the HCAL is made using over a million old, brass shell casings from the Russian Navy back from World War II.

About 34% of the particles produced from LHC pp collisions enter the HE region, so the HE was built to handle high rates (MHz).

The entire HCAL utilizes approximately 70,000 plastic scintillator tiles. The active material

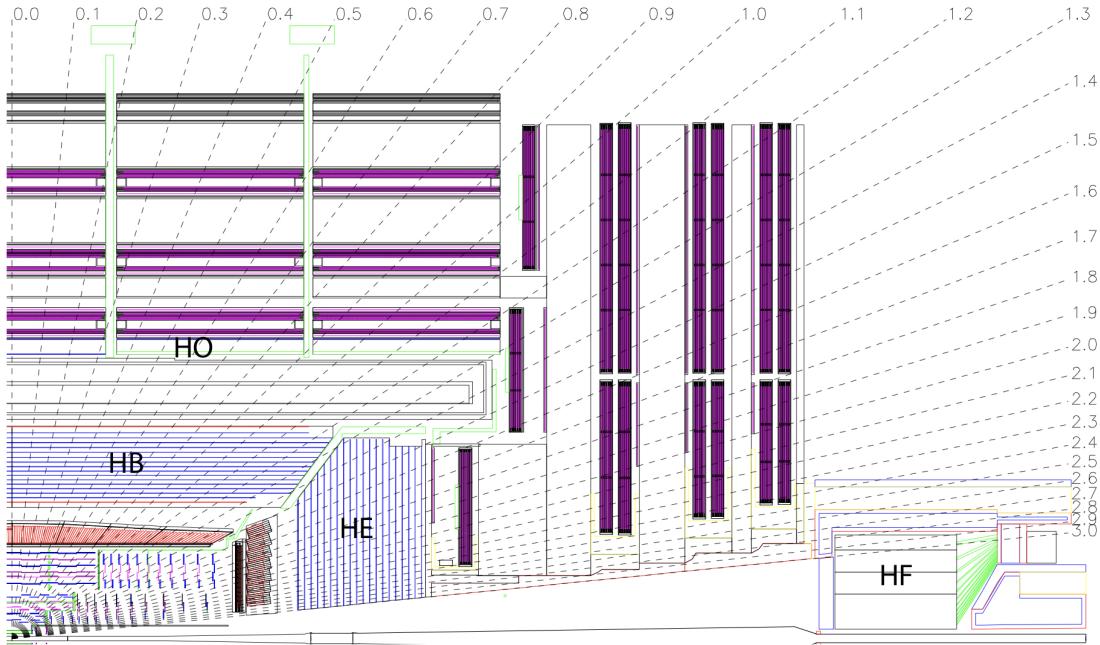


Figure 2-9. A cross-sectional quadrant of CMS showing the locations of the HCAL components: the barrel (HB), outer (HO), endcap (HE), and forward (HF) detectors.

in the HB is 3.7-mm-thick Kuraray SCSN81 plastic scintillator, selected for its radiation hardness and long-term stability. Hadron showers  $\rightarrow$  tiles scintillate  $\rightarrow$  scintillated photons are collected by 0.94-mm-diameter green double-cladded wavelength shifting (WLS) fibers (Kuraray Y-11), which carry the light to hybrid photodiodes (HPD).

### 2.3 The Solenoid and the Steel Return Yoke

The Compact Muon *Solenoid* sports one of the world's most energetic solenoids which is paramount to the success of CMS. Particles that exit the HCAL (subsec. 2.2.2) arrive at the cylindrical magnet which is 12.5 m in length, has a bore diameter of 6 m (6.3 m when cold), and generates a uniform 3.8 T magnetic field parallel to the beam line. To produce such a large and uniform magnetic field inside the approximately  $360 \text{ m}^3$  volume (Fig. 2-10), an 18,000 amp current travels through the 4-layer, superconducting, NbTi coils. This magnetic field is 100,000 times stronger than Earth field at the surface, storing a massive 2.6 GJ of energy—approximately the kinetic energy of an Airbus A320 in flight. The magnet has such a

large stored-energy-to-cold-mass ratio (11.6 KJ/Kg) that it experiences a physical deformation of 0.15% while energizing the field.

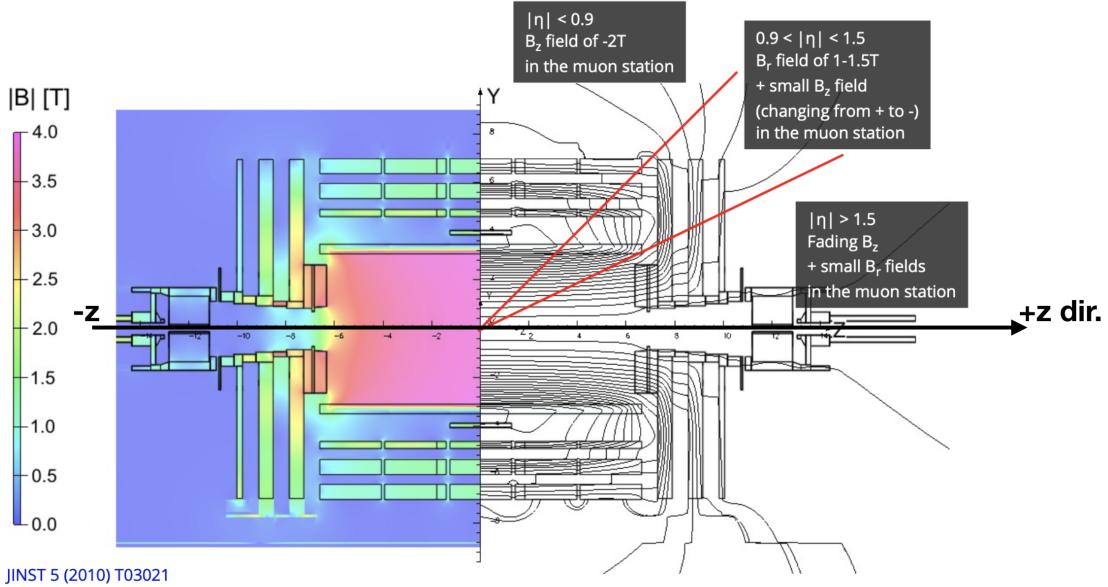


Figure 2-10. A longitudinal cross section of CMS showing the values of the magnetic field over the volume of CMS and various field lines. The magnetic field reaches its maximum of 3.8 T in the center of the detector.

As charged particles travel through any magnetic field, they experience a magnetic (Lorentz) force perpendicular to their direction of travel. The Lorentz force ( $\vec{F}_B$ ) exerted on a particle with charge  $q$  depends on the particle's velocity ( $\vec{v}$ ) and the strength of the magnetic field ( $\vec{B}$ ), given by

$$\vec{F}_B = q\vec{v} \times \vec{B}$$

Since the force is necessarily perpendicular to the velocity, the resulting trajectory is a helix. Projecting the helix on the  $x$ - $y$  plane (since the magnetic field points in the  $+z$  direction) allows the particle tracks to typically be separated from one another. Each track has a corresponding radius of curvature ( $R$ ) which relates to its transverse momentum ( $p_T$ ) through

$$p_T = qBR.$$

The relative change in  $p_T$  (i.e. the momentum resolution) is given by

$$\frac{\delta p_T}{p_T} \propto \frac{p_T}{BL^2}. \quad (2-1)$$

**Steel Return Yoke:** Most of the mass of CMS comes from the *steel return yoke* which helps to redirect the magnetic field back on itself. The yoke system constitutes 10,000 tonnes, which is 89% of the total mass of CMS. It is comprised of 5 wheels and 2 endcaps

# CHAPTER 3

## HIGGS BOSON MASS MEASUREMENT IN THE $H \rightarrow ZZ^* \rightarrow 4\ell$ CHANNEL

### 3.1 Introduction

- Higgs discovery: - EACH OF THESE IS A SENTENCE - higgs discovered in 2012 by CMS, ATLAS. - This was a momentous discovery and often called the “missing puzzle piece” of the SM. - So far it is one of a kind: fundamental scalar particle. - To be certain it is the predicted Higgs boson from the SM, it is necessary to measure its properties and compare them to the predictions of the SM.

- Higgs properties: - Portal to BSM (ref: search for dilepton mass resonance) - There are many results on Higgs properties: spin, charge, decay processes, lifetime, mass. - The last of these is of particular importance: depending on  $m_H$  and  $m_{top}$ , the stability of the Universe.

- ALL previous mass measurements: - Run 1: -  $H \rightarrow 2\gamma$  VALUE -  $H \rightarrow ZZ \rightarrow 4L$  VALUE  
- Run 2: -  $H \rightarrow 2\gamma$  VALUE - (2016)  $H \rightarrow ZZ \rightarrow 4L$  VALUE -  $H \rightarrow bb$  -  $H \rightarrow \mu\mu$  -  $H \rightarrow WW$

- Why this thesis is important: - This thesis describes the methodology and results of the best precision measurement of  $m_H$  to date by using the  $hZZ4l$  decay and Full Run 2 data set from CMS. - Run 2 provides more data -> more precision on measurements of Higgs properties. - In addition to more  $HZZ4l$  events, this analysis provides new techniques, specifically the VX constraint. - Predict  $m_H$  for Run 3, will start soon summer 2022 and provide an approximate 300? /fb of L int. - In 2026(?), HL-LHC provides even more data. ref snowmass paper.

This chapter is structured as follows:

- General overview of the ingredients of the Higgs boson mass measurement (Section 3.2).
- Data sets and simulated samples (Section 3.3).
- Object reconstruction (Section 3.3)
- Event selection (Section 3.3)

SEEMS TO BE A GOOD INTRO. Should it be the intro for the entire thesis?

### 3.2 Analysis Overview

- Want to measure the Higgs boson mass ( $m_H$ ), so need Higgs bosons. - Sift through CMS data for  $H \rightarrow ZZ^* \rightarrow 4\ell$  events (the *signal* process) because S/B ratio is huge: 2. - However

$H \rightarrow ZZ^* \rightarrow 4\ell$  is rare: Although the LHC , the Higgs boson is produced in only 1 out of every billion pp collisions. Even if  $H$  is produced, it will decay into two  $Z$  only a small percentage of the time (2.3%). This percentage is typically expressed as a fraction, called the *branching fraction* or *branching ratio* ( $\mathcal{B}$ ). there is a small probability of only that it will decay into two  $Z$  bosons (2.3%). Those  $Z$  bosons then have only a small probability Furthermore, the Higgs boson has a mean lifetime of only  $1 \times 10^{-23}$  s, so the boson itself will never live long enough to interact directly with the CMS Detector. Therefore, the Higgs boson can only be detected by the daughter particles into which it decays. - By collecting events with the  $4\ell$  final state, we are likely to find signal events. - It's not just the signal process which produces  $4\ell$ : background also makes  $4\ell$  (Section FIXME). - Before analyzing the data, however, it is important to make predictions using simulated samples (Section FIXME). - In order to sort signal from background, use simulated samples

- Form objects from data. - Use objects and conservation of momentum to rebuild parent particles. - The  $Z$  boson has a precisely measured mass of TODO a neutral particle, so the two leptons into which it decays should combine to Group two leptons together, - Form two different pairs of opposite-sign, same-flavor (OSSF) leptons - If it appears that the to select specific hzz4l events (*event selection*). -

### 3.3 Data Sets and Simulated Samples

#### 3.3.1 Data Sets

#### 3.3.2 Simulated Samples

### 3.4 Object Reconstruction

#### 3.4.1 Electrons

#### 3.4.2 Muons

#### 3.4.3 Isolation

#### 3.4.4 Final State Radiation

### 3.5 Event Selection

#### 3.5.1 Triggers

#### 3.5.2 ZZ Candidate Selection

### 3.6 Background estimation

Measurement of the Higgs boson mass requires the accurate modelling of the total event yield in the signal region (SR), into which events can be categorized as either *signal* or *background*. These background events pass the signal event selection (Section 3.5.2) and thus spoil the purity of the signal events. This introduces further uncertainty into the final Higgs boson mass measurement. Therefore, it is a priority to both reduce and predict the expected number of background events, which can be split into *irreducible backgrounds* (IB) and *reducible background* (RB) processes.

#### 3.6.1 Irreducible background

IB processes produce two Z bosons and each Z subsequently decays into two prompt leptons (leptons that emerge directly from the primary vertex). This reliably produces a  $4\ell$  final state, whose prompt leptons typically get reconstructed as four leptons that pass tight selection (*PTS leptons*), as defined in Sections 3.4.1 and 3.4.2. The IB event then looks indistinguishable from the 4 PTS lepton of the *signal* process and cannot be reduced; Thus, throwing away IB process could mean throwing away a signal event. Since IB event cannot be reduced, it's called irreducible background. The two IBs for the HZZ4L analysis are:

- $gg \rightarrow ZZ \rightarrow 4\ell$  (gluon-gluon fusion),
- $q\bar{q}/gg \rightarrow ZZ \rightarrow 4\ell$  (quark-antiquark annihilation).

### 3.6.2 Reducible background

While irreducible background processes produce 4 prompt leptons, RB processes produce varying numbers of prompt and nonprompt leptons. Nonprompt leptons emerge from 3 main sources:

- misidentifying light-flavor hadrons (e.g.,  $\pi^\pm$ ) as leptons,
- heavy-flavor hadrons that decay mid-flight into leptons,
- and the asymmetric conversion of photons into electrons.

This analysis is tailored to efficiently reconstruct prompt leptons as those which pass tight selection (*PTS leptons*), whereas nonprompt leptons typically fail tight selection (*FTS leptons*). Tight and loose lepton selections are defined in Sections 3.4.1 and 3.4.2. Ideally, the SR would contain only the events with 4 prompt leptons, which would always be tagged as 4 PTS leptons. However, sometimes a truly nonprompt lepton is misidentified as a PTS lepton (sometimes called a *fake lepton*), depending on the kinematical properties of the lepton. This misidentification rate ( $f$ , sometimes called the *fake rate*) is due to imperfect detector performance, inefficiencies in reconstruction, and the specific choice of lepton selections used in the analysis.

If an event produces prompt and nonprompt leptons but is reconstructed as 4 PTS leptons (a 4P event), then it contaminates the SR. These processes constitute the RB processes (sometimes called “ $Z + X$ ”). Examples of RB processes for this analysis include:

- $Z + \text{jets}$  (yields 2 prompt leptons)
- $t\bar{t} + \text{jets}$  (yields 2 prompt leptons)
- $WZ + \text{jets}$  (yields 3 prompt leptons)
- $Z(Z/\gamma^*) + \text{jets}$  (yields 4 prompt leptons).

The careful estimation of these RB contributions to the 4P region is necessary for the precise measurement of the Higgs boson mass.

### 3.6.2.1 OS Method

The goal of the OS Method is to estimate the number of opposite-sign same-flavor (OSSF)  $4\ell$  events produced by RB process that “contaminate” the 4P region ( $N_{4P}^{RB}$ ). However, the typical approach of using simulated samples does not model RB well, since RB processes (e.g., Z + jets) rely on higher-order effects like jet modelling which are not yet accurately simulated. Instead, a data-driven approach is used.

The logic of the OS Method is to study events in data that are similar to, but not exactly the same as, those found in the 4P region. Thus, events in data are sorted into 2 control regions (CRs), both of which are orthogonal to the 4P region and to each other:

- the 3P1F CR (built from 3 PTS and 1 FTS leptons)
- the 2P2F CR (built from 2 PTS and 2 FTS leptons).

The event selection for the 2P2F and 3P1F CRs is almost identical to that of the SR (Section 3.5.2), except that the FTS lepton(s) must go into building the  $Z_2$ . The events which contribute most to the 2P2F CR are those that produce 2 prompt and 2 nonprompt leptons (*2pr events*). Events that produce 3 prompt and 1 nonprompt leptons (*3pr events*) contribute mostly to the 3P1F CR.

To understand the logic behind the final formula for  $N_{4P}^{RB}$  (Eq. 3-13), it is helpful to first analyze a single event: suppose event  $k$  is a 2pr event which contributes to the 2P2F CR a certain event weight ( $w_{2\text{pr} \rightarrow 2\text{P2F}}^k$ ). This event weight is built from the product of analysis weights (pileup, L1 pre-firing, etc., whose product will be called  $\hat{w}^k$ ) and the reconstruction efficiencies ( $\epsilon$ ) of each lepton ( $\ell_n$ ):

$$w_{2\text{pr} \rightarrow 2\text{P2F}}^k = \hat{w}^k \cdot \epsilon_P^{\text{pr}}(\ell_1^k) \cdot \epsilon_P^{\text{pr}}(\ell_2^k) \cdot \epsilon_F^{\text{np}}(\ell_3^k) \cdot \epsilon_F^{\text{np}}(\ell_4^k), \quad (3-1)$$

where the superscript of  $\epsilon$  indicates the lepton promptness (pr = prompt, np = nonprompt), the subscript indicates the lepton tightness status (P = PTS, F = FTS). To simplify the equations that follow,  $\hat{w}^k$  will be set to unity. If the reconstruction efficiencies are the same for all leptons of a particular category (e.g.  $\epsilon_P^{\text{pr}}$ ) across all events, then Eq. 3-1 reduces to:

$$w_{2\text{pr} \rightarrow 2\text{P2F}}^k = (\epsilon_P^{\text{pr}})^2 (\epsilon_F^{\text{np}})^2. \quad (3-2)$$

Although a 2pr event mostly contributes to the 2P2F CR, a nonprompt lepton may be misidentified as a PTS lepton depending on  $\epsilon_P^{np}$ . In this case, event  $k$  would then fall into the 3P1F CR and, being careful to allow only one nonprompt PTS lepton at a time, contributes an effective weight of

$$\begin{aligned} w_{2\text{pr} \rightarrow 3\text{P1F}}^k &= (\epsilon_P^{\text{pr}})^2 [\epsilon_P^{np}(\ell_1^k) \cdot \epsilon_F^{np}(\ell_2^k) + \epsilon_F^{np}(\ell_1^k) \cdot \epsilon_P^{np}(\ell_2^k)] \\ &= (\epsilon_P^{\text{pr}})^2 [2\epsilon_P^{np}\epsilon_F^{np}]. \end{aligned} \quad (3-3)$$

Using the fact that a (non)prompt lepton is exclusively either PTS or FTS,

$$\epsilon_P^{\text{pr(np)}} + \epsilon_F^{\text{pr(np)}} = 1,$$

recognizing that  $\epsilon_P^{np} \equiv f$  (Section 3.6.2), and defining  $\epsilon_P^{\text{pr}} \equiv \epsilon$ , then Eq. 3-3 becomes

$$w_{2\text{pr} \rightarrow 3\text{P1F}}^k = 2\epsilon^2 f(1-f). \quad (3-4)$$

The prediction of Eq. 3-4 can be seen in Figures 3-4–3-6 for 2016–2018 UL data.

A smaller fraction of the time, both nonprompt leptons from a 2pr event may be misidentified as PTS leptons. In this case, event  $k$  contributes to the 4P region an effective weight of

$$w_{2\text{pr} \rightarrow 4\text{P}}^k = \epsilon^2 f^2, \quad (3-5)$$

where it is assumed that both leptons have the same misidentification rate. Similar equations can be derived for the contributions of a 3pr event to the 3P1F CR and to the 4P region:

$$w_{3\text{pr} \rightarrow 3\text{P1F}}^k = \epsilon^3 (1-f) \quad (3-6)$$

$$w_{3\text{pr} \rightarrow 4\text{P}}^k = \epsilon^3 f. \quad (3-7)$$

Since a 3pr event needs only 1 nonprompt PTS lepton to be included in the 4P region (therefore, carrying only 1 factor of  $f$ ), a 3pr event tends to contribute more weight to 4P than does a 2pr event (which carries  $f^2$ ). If the total number of 2pr and 3pr events is  $X_2$  and  $X_3$ , respectively, then

the expected number of 4P events using Eqs. 3-5 and 3-7 is

$$\begin{aligned}
N_{4P}^{RB} &= \sum_{k=1}^{X_2} w_{2pr \rightarrow 4P}^k + \sum_{m=1}^{X_3} w_{3pr \rightarrow 4P}^m \\
&= \sum_{k=1}^{X_2} \epsilon^2 f^2 + \sum_{m=1}^{X_3} \epsilon^3 f \\
&= f^2 \epsilon^2 X_2 + f \epsilon^3 X_3.
\end{aligned} \tag{3-8}$$

Only the quantities  $f$  and  $\epsilon$  are known, so  $X_2$  and  $X_3$  must be estimated. This is achieved by relating  $X_2$  to the observed number of 2P2F events ( $N_{2P2F}$ ). Using Eq. 3-2 across all 2pr events gives:

$$\begin{aligned}
N_{2P2F} &= \sum_{k=1}^{X_2} w_{2pr \rightarrow 2P2F}^k \\
&= \sum_{k=1}^{X_2} \epsilon^2 (1-f)^2 \\
&= (1-f)^2 \epsilon^2 X_2
\end{aligned} \tag{3-9}$$

The strategy to relate  $X_3$  to the observed number of 3P1F events ( $N_{3P1F}$ ) is not as straightforward as relating  $X_2$  to  $N_{2P2F}$ , since two other sources also contribute to the 3P1F CR:

- A 2pr RB process can yield one nonprompt PTS lepton (via Eq. 3-4).
- A 4-prompt IB process ( $q\bar{q}/gg \rightarrow ZZ \rightarrow 4\ell$ , “ZZ”) can yield one FTS lepton.

The second of these is well estimated from simulation, since ZZ produces 4 prompt leptons.

Supposing that the total number of simulated ZZ events is  $X_4^{ZZ}$ , then event  $k$  belonging to these events contributes to the 3P1F CR an effective weight of

$$w_{4prZZ \rightarrow 3P1F}^k = 4\epsilon^3 \epsilon_F^{pr}, \tag{3-10}$$

which accounts for any of the 4 prompt leptons to be reconstructed as a FTS lepton while the

others are PTS leptons. Incorporating Eqs. 3-4, 3-6, and 3-10 for all events relates  $X_3$  to  $N_{3\text{P1F}}$ :

$$\begin{aligned}
N_{3\text{P1F}} &= \sum_{j=1}^{X_3} w_{3\text{pr} \rightarrow 3\text{P1F}}^j + \sum_{k=1}^{X_2} w_{2\text{pr} \rightarrow 3\text{P1F}}^k + \sum_{m=1}^{X_4^{\text{ZZ}}} w_{4\text{prZZ} \rightarrow 3\text{P1F}}^m \\
&= \sum_{j=1}^{X_3} \epsilon^3 (1-f) + \sum_{k=1}^{X_2} 2\epsilon^2 f (1-f) + \sum_{m=1}^{X_4^{\text{ZZ}}} 4\epsilon^3 \epsilon_{\text{F}}^{\text{pr}} \\
&= (1-f)\epsilon^3 X_3 + 2f(1-f)\epsilon^2 X_2 + 4\epsilon^3 \epsilon_{\text{F}}^{\text{pr}} X_4^{\text{ZZ}} \\
&= (1-f)\epsilon^3 X_3 + 2f(1-f)\epsilon^2 X_2 + N_{3\text{P1F}}^{\text{ZZ}}, \tag{3-11}
\end{aligned}$$

where  $N_{3\text{P1F}}^{\text{ZZ}}$  is simply the raw (integer) number of ZZ events that pass 3P1F selections, obtained directly from simulation.

At this point,  $N_{4\text{P}}^{\text{RB}}$  can be isolated by combining Eqs. 3-8, 3-9, and 3-11:

$$N_{4\text{P}}^{\text{RB}} = \left( \frac{f}{1-f} \right) N_{3\text{P1F}}^{\text{Data}} - \left( \frac{f}{1-f} \right)^2 N_{2\text{P2F}}^{\text{Data}} - \left( \frac{f}{1-f} \right) N_{3\text{P1F}}^{\text{ZZ}}. \tag{3-12}$$

Thus, Eq. 3-12 estimates the RB contribution to the 4P region by using a single lepton misidentification rate to reweight the raw yields of events found in the 3P1F and 2P2F CRs of data, and also the 3P1F CR of ZZ.

It should be mentioned that the above formula was derived assuming that the per event analysis weights were set to unity ( $\hat{w}^k = 1$ ) before being scaled by the misidentification rates. It was also assumed that  $f$  is a constant, for all nonprompt leptons misidentified as PTS leptons across all events, which is not the case as is shown in Figure 3-7. Thus, an extension of Eq. 3-12 can be formed by assigning misidentification rates that depend on the kinematical variables per lepton, by restoring the analysis weights per event, and by summing over the total number of raw yields per CR:

$$N_{4\text{P}}^{\text{RB}} = \sum_{i=1}^{N_{3\text{P1F}}^{\text{Data}}} \hat{w}^i \frac{f_i}{1-f_i} - \sum_{j=1}^{N_{2\text{P2F}}^{\text{Data}}} \hat{w}^j \frac{f_{1,j}}{1-f_{1,j}} \frac{f_{2,j}}{1-f_{2,j}} - \sum_{k=1}^{N_{3\text{P1F}}^{\text{ZZ}}} \hat{w}_{\text{ZZ}}^k \frac{f_k}{1-f_k}, \tag{3-13}$$

where  $f_{1,j}$  and  $f_{2,j}$  are the misidentification rates of the first and second FTS leptons, respectively, found in the  $j^{\text{th}}$  2P2F event and  $\hat{w}_{\text{ZZ}}^k$  accounts for the differential QCD and electroweak  $k$ -factors

defined in Section 3.6.1.

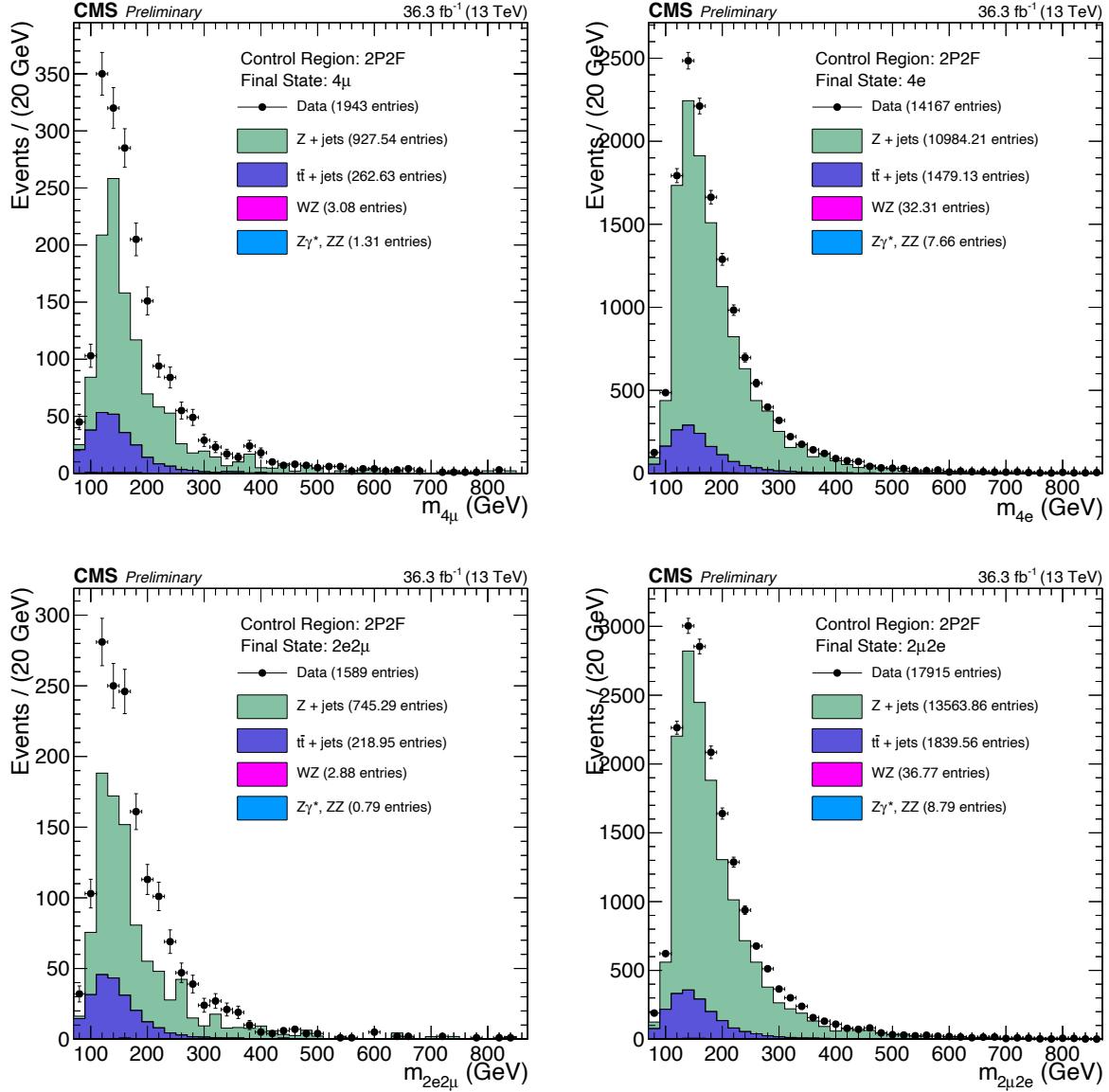


Figure 3-1. Events from 2016 UL data that pass 2P2F CR selections (black markers) are compared to the stacked 2P2F distributions of simulated samples ( $Z + \text{jets}$ ,  $t\bar{t} + \text{jets}$ ,  $WZ$ ,  $Z/\gamma^*$ ,  $ZZ$ ). The results are split into the  $4\ell$  final states:  $4\mu$  (top left),  $4e$  (top right),  $2e2\mu$  (bottom left),  $2\mu2e$  (bottom right).

### 3.6.2.2 Lepton misidentification rate measurement

As mentioned in Section 3.6.2, the lepton misidentification rate ( $f$ ) is the probability that a nonprompt lepton will pass tight selections (PTS). The value  $f$  is a function of the flavor

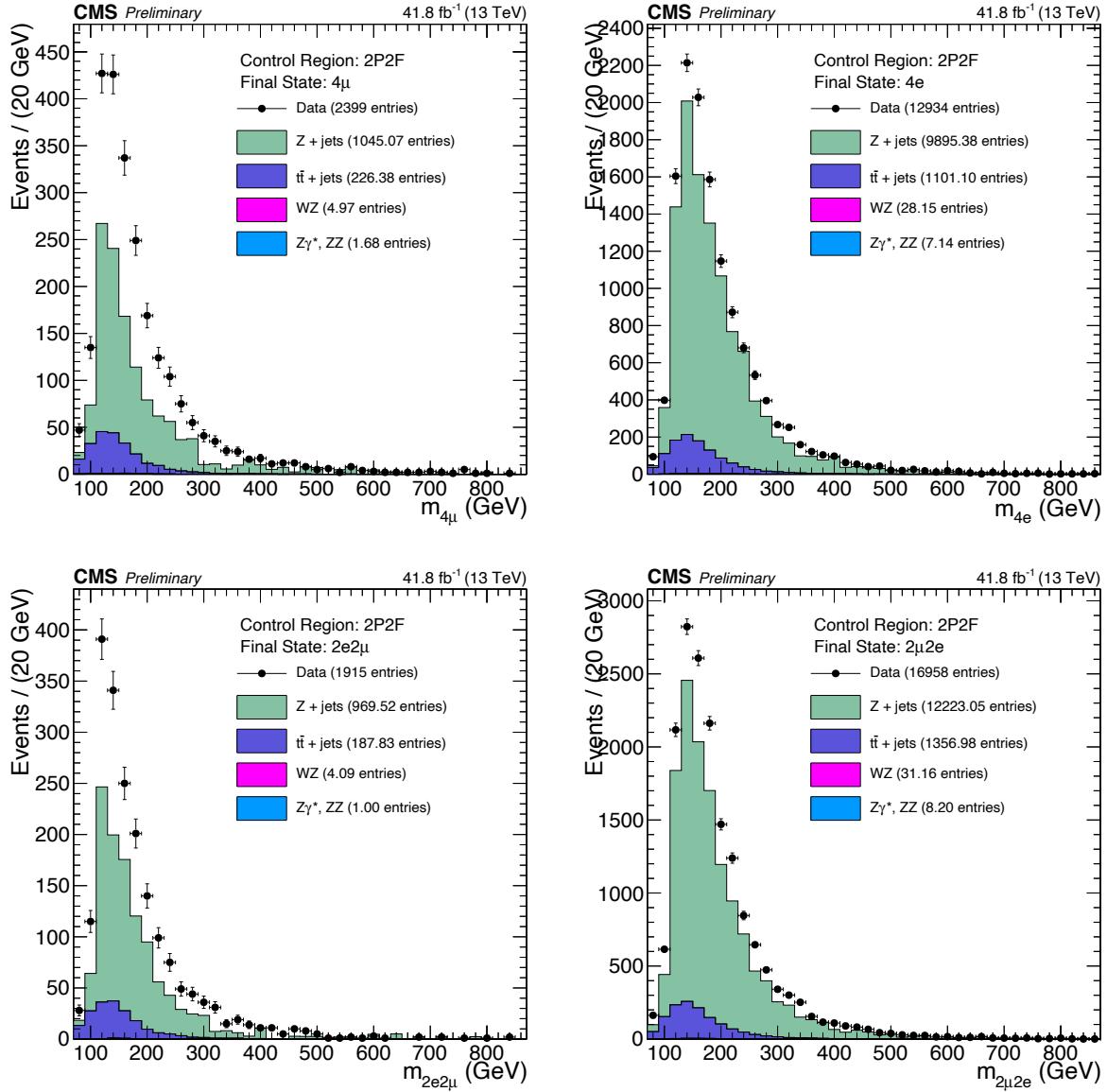


Figure 3-2. Events from 2017 UL data that pass 2P2F CR selections (black markers) are compared to the stacked 2P2F distributions of simulated samples ( $Z + \text{jets}$ ,  $t\bar{t} + \text{jets}$ , WZ,  $Z/\gamma^*$ , ZZ). The results are split into the  $4\ell$  final states:  $4\mu$  (top left),  $4e$  (top right),  $2e2\mu$  (bottom left),  $2\mu2e$  (bottom right).

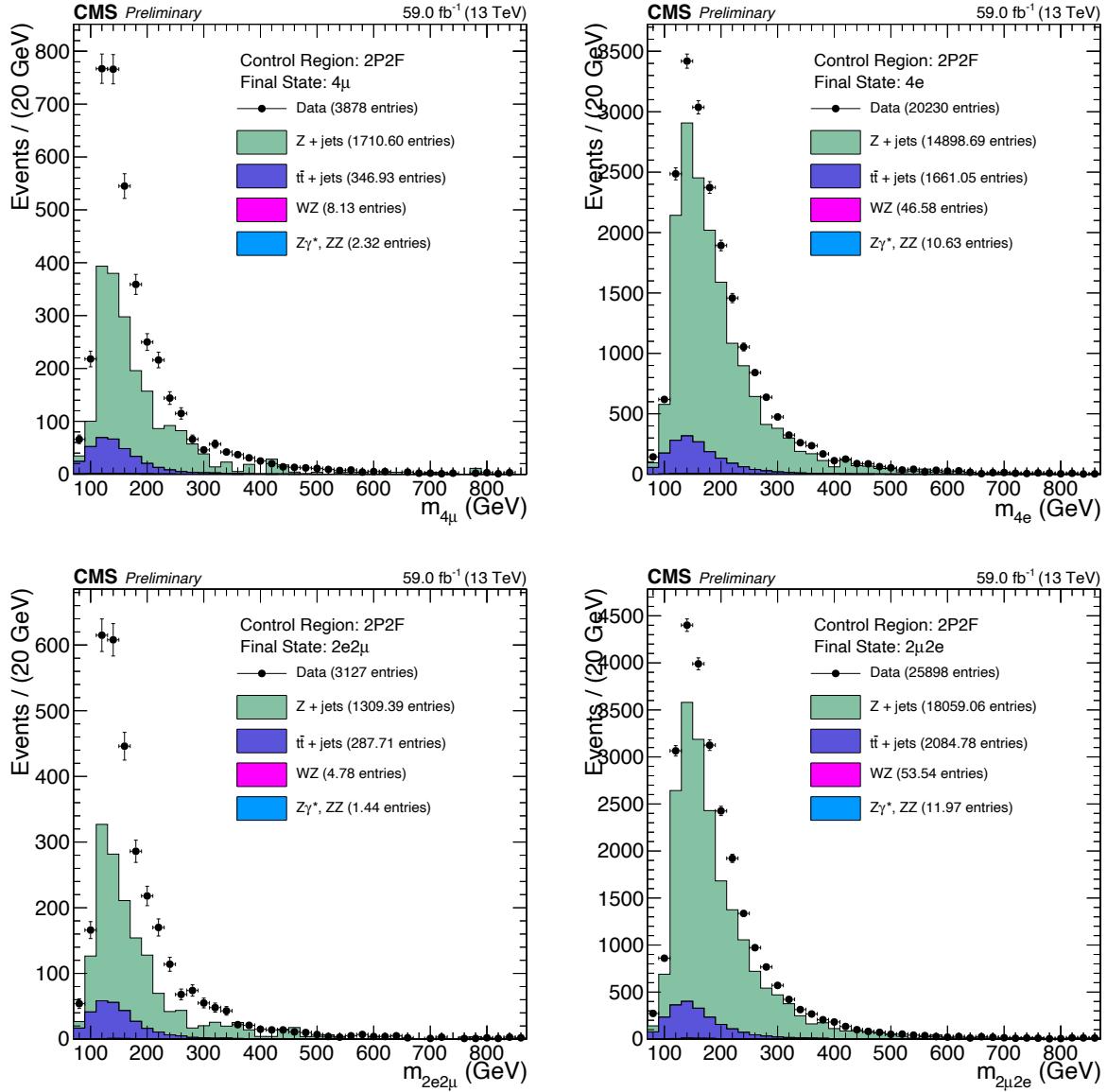


Figure 3-3. Events from 2018 UL data that pass 2P2F CR selections (black markers) are compared to the stacked 2P2F distributions of simulated samples ( $Z + \text{jets}$ ,  $t\bar{t} + \text{jets}$ , WZ,  $Z/\gamma^*$ , ZZ). The results are split into the  $4\ell$  final states:  $4\mu$  (top left),  $4e$  (top right),  $2e2\mu$  (bottom left),  $2\mu2e$  (bottom right).

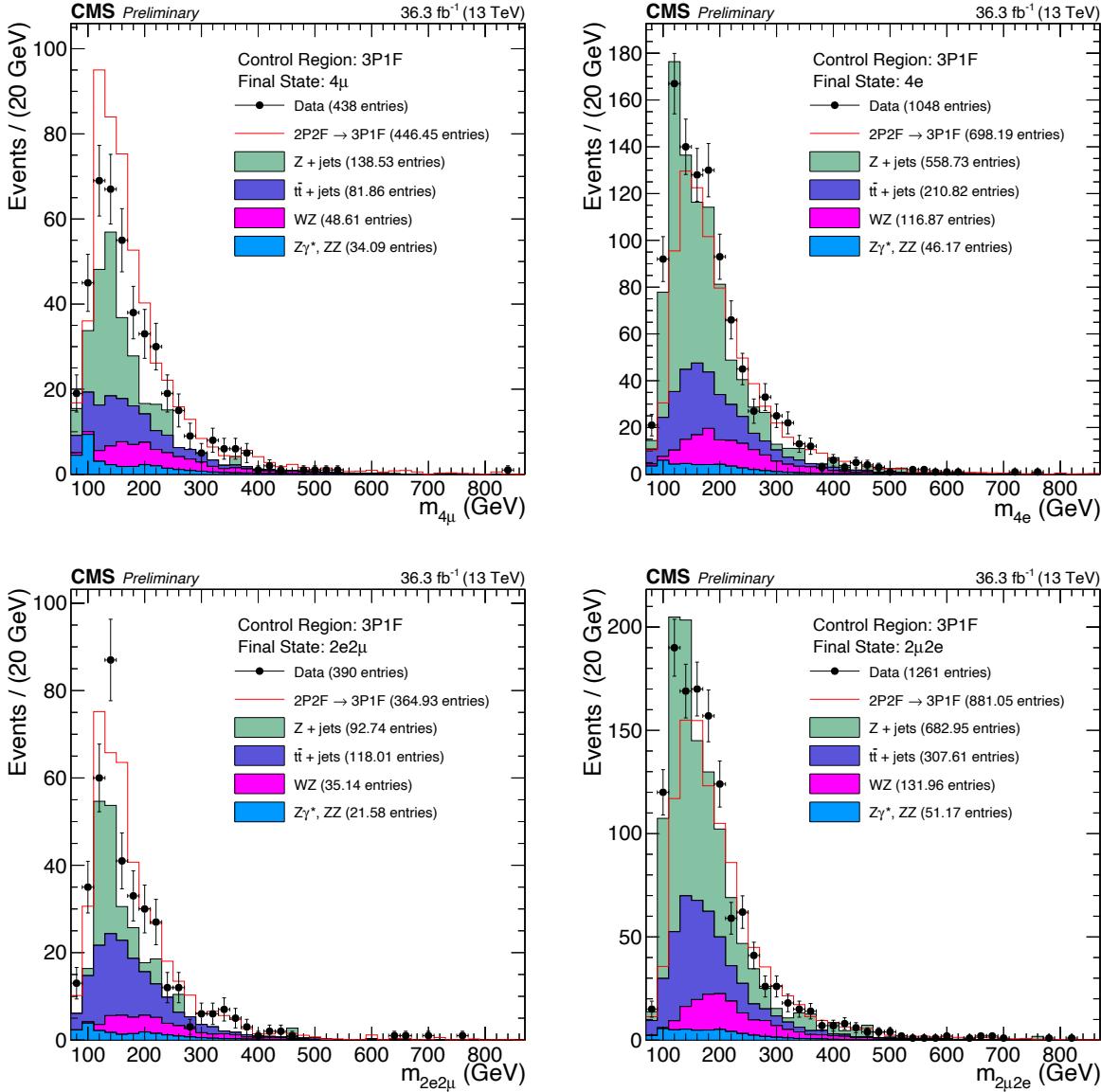


Figure 3-4. Events from 2016 UL data that pass 3P1F CR selections (black markers) are compared to the stacked 3P1F distributions of simulated samples ( $Z + \text{jets}$ ,  $t\bar{t} + \text{jets}$ , WZ,  $Z/\gamma^*$ , ZZ) and to the predicted contribution of 2-prompt-2-nonprompt-lepton processes to the 3P1F CR (red line). This prediction is obtained by making a distribution of all event weights given by Eq. 3-4 and stacking that on top of the WZ and ZZ distributions. The results are split into the  $4\ell$  final states:  $4\mu$  (top left),  $4e$  (top right),  $2e2\mu$  (bottom left),  $2\mu2e$  (bottom right).

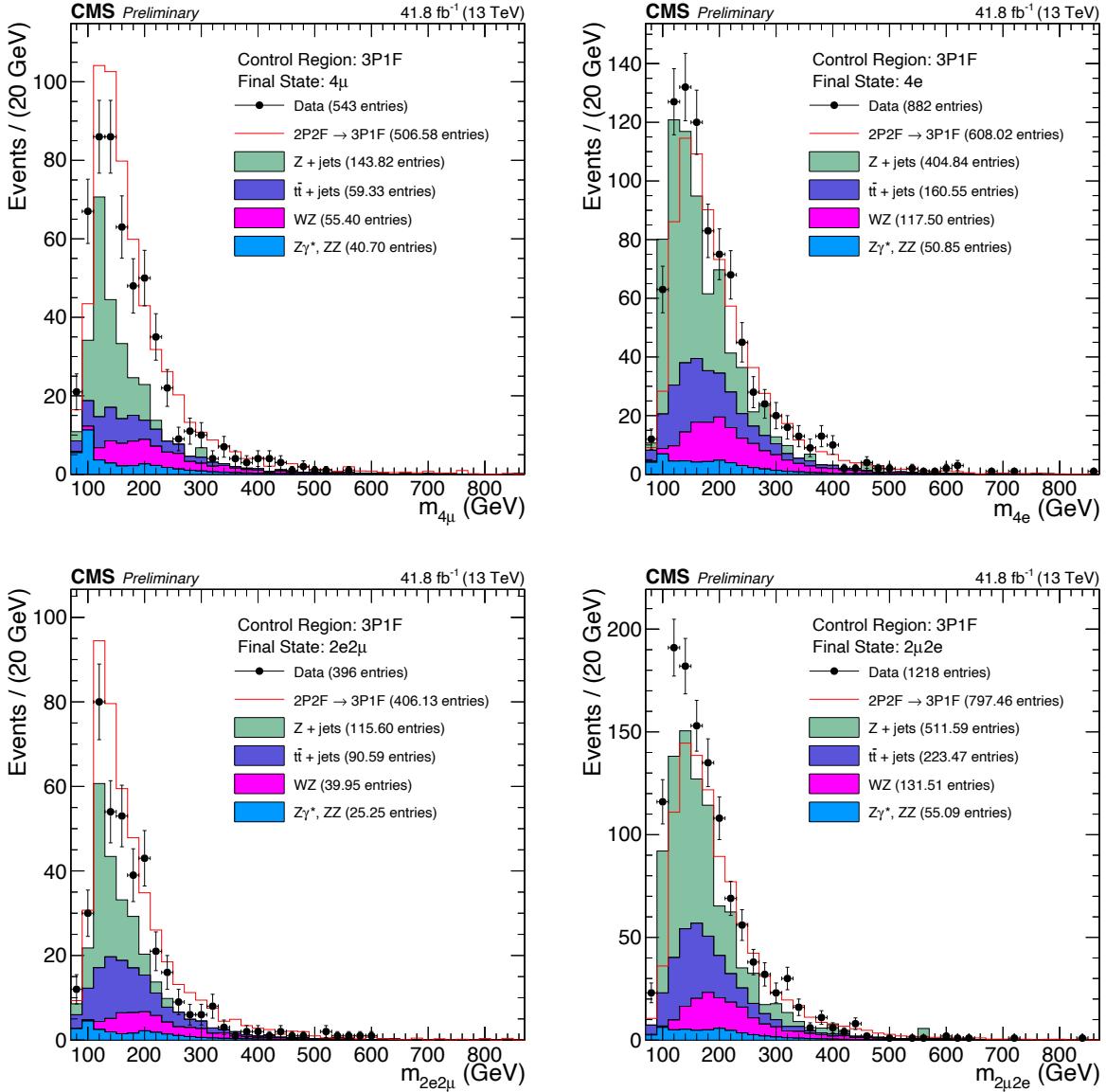


Figure 3-5. Events from 2017 UL data that pass 3P1F CR selections (black markers) are compared to the stacked 3P1F distributions of simulated samples (Z + jets, tt + jets, WZ, Z $\gamma^*$ , ZZ) and to the predicted contribution of 2-prompt-2-nonprompt-lepton processes to the 3P1F CR (red line). This prediction is obtained by making a distribution of all event weights given by Eq. 3-4 and stacking that on top of the WZ and ZZ distributions. The results are split into the 4 $\ell$  final states: 4 $\mu$  (top left), 4e (top right), 2e2 $\mu$  (bottom left), 2 $\mu$ 2e (bottom right).

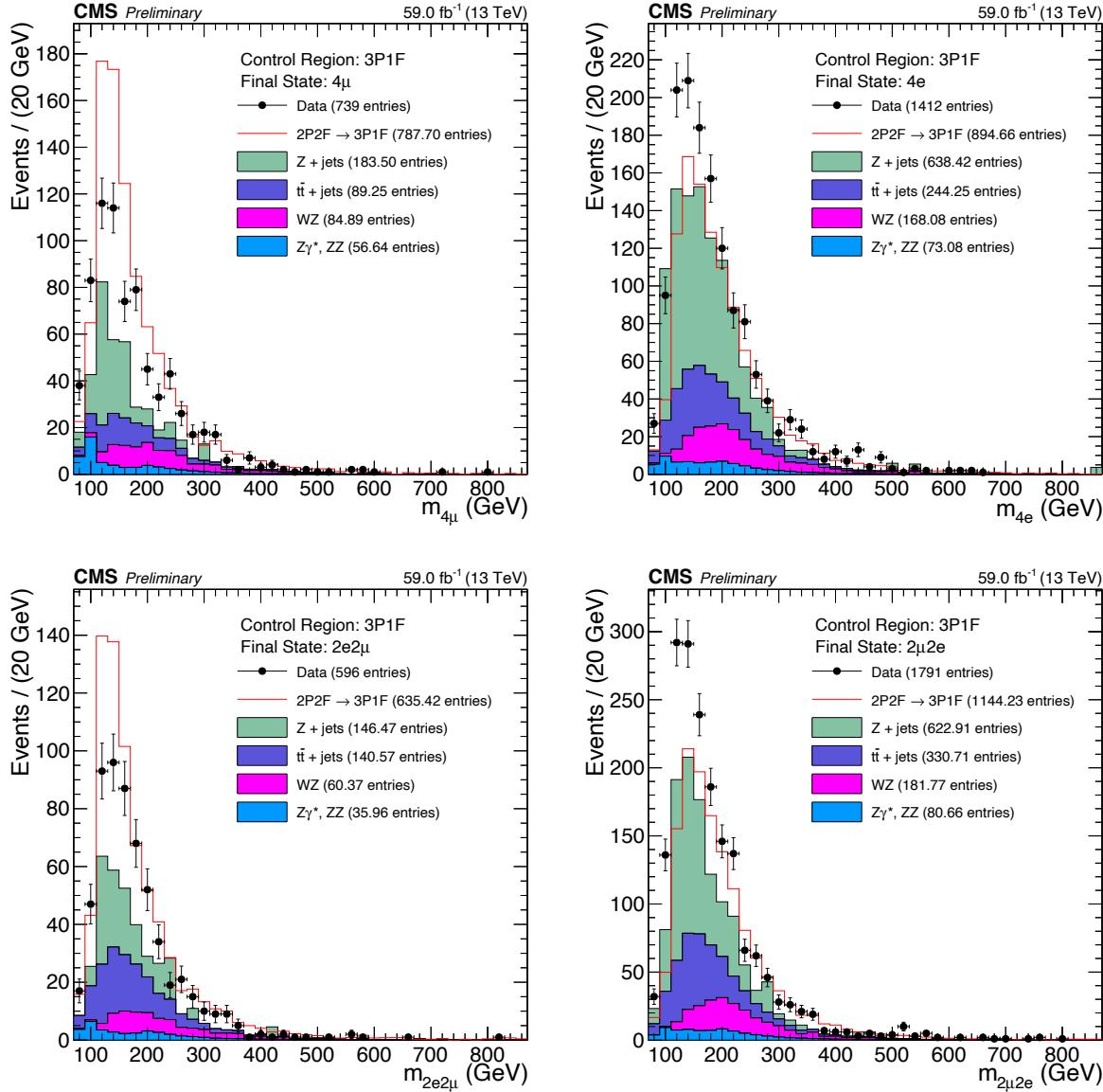


Figure 3-6. Events from 2018 UL data that pass 3P1F CR selections (black markers) are compared to the stacked 3P1F distributions of simulated samples ( $Z + \text{jets}$ ,  $t\bar{t} + \text{jets}$ ,  $WZ$ ,  $Z/\gamma^*$ ,  $ZZ$ ) and to the predicted contribution of 2-prompt-2-nonprompt-lepton processes to the 3P1F CR (red line). This prediction is obtained by making a distribution of all event weights given by Eq. 3-4 and stacking that on top of the  $WZ$  and  $ZZ$  distributions. The results are split into the  $4\ell$  final states:  $4\mu$  (top left),  $4e$  (top right),  $2e2\mu$  (bottom left),  $2\mu2e$  (bottom right).

$(\ell = e, \mu)$ ,  $p_T$ , and  $\eta$  of a lepton. The misidentification rate is calculated by simply counting the number of nonprompt PTS leptons ( $N_P^{\text{np}}$ ) that enter a particular  $\ell, p_T, |\eta|$  bin compared to the total number of loose probe leptons ( $N_L^{\text{np}}$ ) in the same bin:

$$f(\ell, |\eta|, p_T) = \frac{N_P^{\text{np}}}{N_L^{\text{np}}}. \quad (3-14)$$

The  $p_T^e$  bin edges are [5–10–20–30–40–50–80] GeV and the  $p_T^\mu$  bin edges are [5–7–10–20–30–40–50–80] GeV. The nonprompt leptons used to measure  $f$  are taken from events in data with a signature like that of  $Z + \ell_L$ , where  $Z$  is a  $Z$  boson and  $\ell_L$  is a loose lepton. By construction, this region of events is orthogonal to the 2P2F, 3P1F, and 4P regions, and provides a clean source of  $\ell_L$ . The loose lepton, whose selection is defined in Sections 3.4.1 and 3.4.2, is also called the *probe* lepton. The probe lepton is either a PTS or FTS lepton and is counted towards both the numerator and denominator of Eq. 3-14.

Events are selected that satisfy the following criteria:

- The event has exactly 3 leptons.
- The event contains  $E_T^{\text{miss}} < 25$  GeV.
- Two of the leptons form a  $Z$  candidate. A  $Z$  candidate is formed when:
  - The lepton pair is OSSF.
  - Both leptons PTS.
  - The leading lepton has  $p_T > 20$  GeV.
  - The subleading lepton has  $p_T > 10$  GeV.
  - The lepton pair satisfies  $|m_{\ell\ell} - m_{Z_{\text{PDG}}}| < 7$  GeV
- The third and final lepton is loose (and may be a PTS or FTS lepton).
- Suppress QCD processes: probe lepton and OS lepton from  $Z$  have  $m_{\ell\ell} > 4$  GeV.

The calculation of  $f$  requires that  $\ell_L$  is a nonprompt lepton but since data events were used, this may not be the case. For example, the decay of  $WZ$  produces 3 prompt leptons and so this contribution must be subtracted. Thus, the number of expected prompt probe leptons from  $WZ$  events is subtracted from both the numerator and denominator in Eq. 3-14 for each  $\ell, p_T, |\eta|$  bin.

The final OS Method misidentification rates for electrons and muons are shown in Figure 3-7 using 2016–2018 UL data.

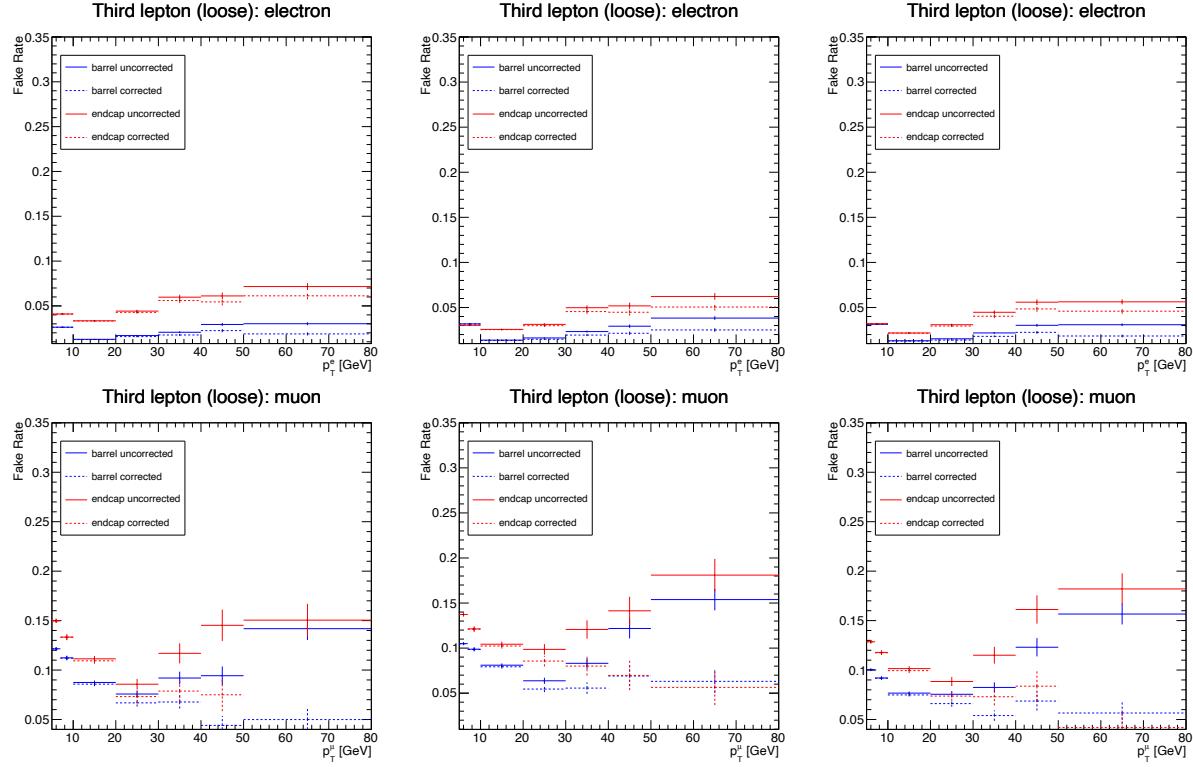


Figure 3-7. Electron (top row) and muon (bottom row) fake rates evaluated using the OS Method vs. the  $p_T$  of the probe lepton for 2016, 2017, and 2018 UL data (left, middle, right columns, respectively).

## CHAPTER 4 SUMMARY

words.

## REFERENCES

- [1] CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3 (8):S08004, August 2008. doi: 10.1088/1748-0221/3/08/S08004. URL <https://ui.adsabs.harvard.edu/abs/2008JInst...3S8004C>.